HIGH LEVEL DOCUMENT

(HLD)

Project Title : **Thyroid DiseaseDetection**

Technologies : Machine Learning

Technology : Domain Healthcare

Project Difficulties level : Intermediate

.Document Version Control

Date :

Version : 1.0

Description : Initial

Author : Himanshu

.Contents

1 Introduction

1.1 Why this High-Level Design Document?

1.2 Scope

1.3 How this HLD will serve?

1.4 Definations

2 General Description

2.1 Product Perspective

2.2 Problem Statement

2.3 Proposed Solution

2.4 Further Improvements

2.5 Tech Requirements

2.6 Data Requirements

2.7 Tools Used

2.8 Constraints

2.9 Assumptions

3 Design Details

3.1 Process Flow

3.1.1 Model Training and Evaluation

3.1.2 Deployment Process

3.2 Event Log

3.3 Error Handling

3.4 Reusability

3.5 Application Compatibility

3.6 Deployment

4 Conclusion

1 Introduction

1.1 Why High-Level Design Document?

As a machine learning practitioner, it is easy to overlook some of the many difficulties associated with building machine learning models while starting any particular project. Could the data be biased? What metric performance would be enough? Am I solving the right problem in the first place? Will the project be ready in time?

-----Training and building the Model is often the easiest part of the ML project. The main difficulty can be something else, for example,

>> Am I asking the correct questions to the data?

>> Do I've enough data to answer the

questtions?

>> Am I solving the right problem statement altogether?

>> Do I really have the clarity about the problem statement? Etc.

In worst senario, all this can lead to a creation of a Model which will never see the production server. Luckily, we can prevent this unfortunate scenario by creating a design doc or HLD at the first place.

1.2 Scope

Design documents can bring the following benefits to machine learning projects:

1. It forces the author to organize their thoughts and get feedback on their ideas.

2. A design doc greatly improves communication with other teams.

3. It clarifies the scope of the project, highlights expectations and risks, and makes sure all the important questions about the project get answered.

1.3 How this HLD will serve?

This HLD document is intended to help detect ontradictions prior to coding and can be used as areference manual for how the modules interact at a high level.

This HLD is going to serve the following:

1. All the design aspects
2. Design of the User Interface
3. Software and tools required to achive the objective
4. Data Source
5. Performance requirements
6. Design Features and architecture of the project
7. Reliability
8. Maintainance or Model training Frequency
9. Future Scopes
10. Limitations

1.4 Definations

Term Description

T D D : Thyroid Disease Detection

IDE : Integrated Development Environment

AWS : Amazon Web Services

2 General Description

2.1 Product Perspective

The TDD is a ML based Thyroid disease Prediction system that predicts thyroid disease detection using suitable ML Algorithm using historical data. This can be helpful to give an approximate idea of how

much disease the patient has or not .

2.2 Problem Statement

Thyroid disease is one of the most common disease with endocrine disorder in the human population today. For example hyperthyroidism (over) and hypothyroidism (under), which are relate to release of amount of thyroid hormones the thyroid gland produces and whether it is over active trusted source (when thyroid gland makes too much thyroid hormone) or under active trusted source (when the thyroid gland doesn't make enough thyroid hormone). **We need to identify whether the patient has thyroid or not.**

2.3 Proposed Solution

We need to

2.4 Further Improvements

2.5 Tech Requirements

i) AWS paid services for deployment

ii) Flawless Database

iii) Github action and Workflow services

2.6 Data Requirements

Updated data from all airline services with few mandatory features like Source, Destination, Airline Name, Duration, Route or No of Stops etc.

2.7 Tools Used

i) VS Code as IDE

ii) MongoDB as Database

iii) Jupyter Notebook to perform all EDA, FE, Model Building and selecting the best model.

iv) Pandas and Numpy for Data Analysis

v) Python programming language

vi) Seaborn and Matplotlib as visualization tools

viii) Github as version control system

ix) Github action for CI/CD

x) AWS Elastic Beanstalk and S3 Bucket for Deployment

xi) Flask

xii) Scikit-learn for Model building and Evaluation

2.8 Constraints

Accuracy of the prediction totally depends on continuous availability of updated data and retraining of the model with updated data. The predicted disease is a approximate amount which may vary from the current disease.

In real world the Thyroid disease depending on various factors on which our model is not trained.

2.9 Assumptions

3 Design Details

3.1 Process Flow

There will two main features -

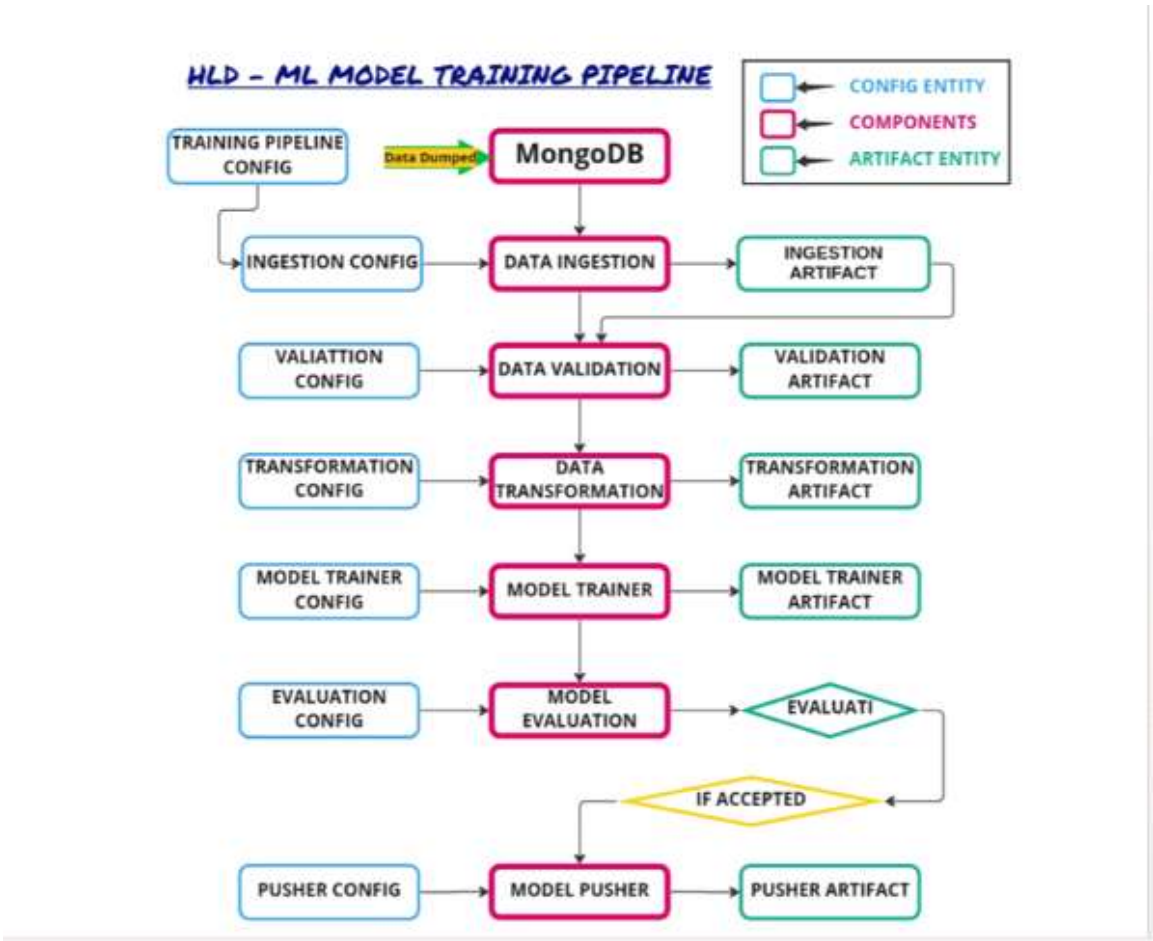
I. Model Training Pipeline and

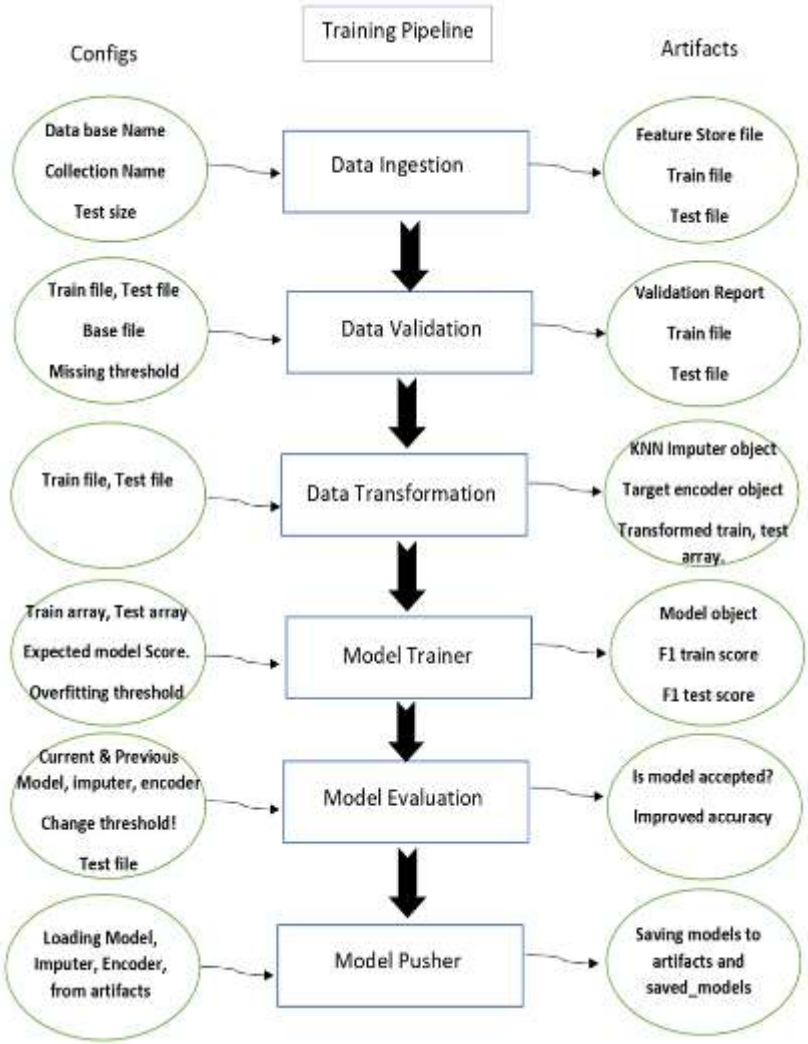
II. Prediction.

Following Diagram represents the proposed process flow of both features.

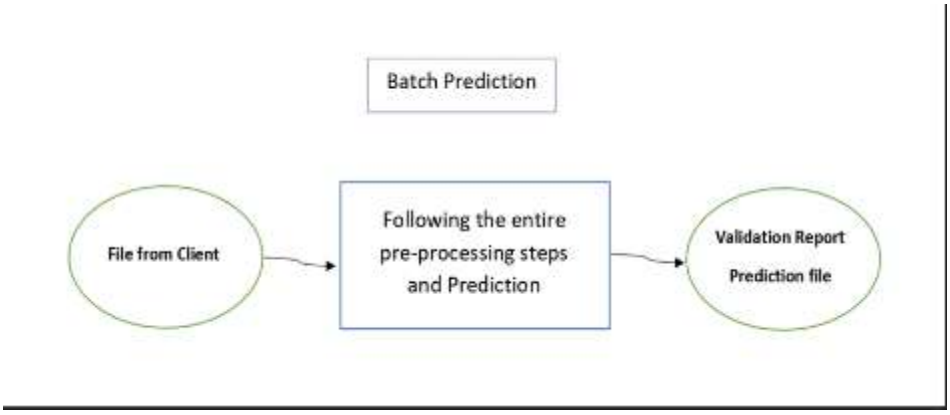
3.1.1 Model Training Pipeline

To be initiated manually when new data is available in database.


HLD – ML MODEL TRAINING PIPELINE
CONFIG ENTITY
COMPONENTS
ARTIFACT ENTITY
TRAINING PIPELINE CONFIG
Data Dumped
MongoDB
INGESTION CONFIG
DATA INGESTION
INGESTION ARTIFACT
VALIATTION CONFIG
DATA VALIDATION
VALIDATION ARTIFACT
TRANSFORMATION CONFIG
DATA TRANSFORMATION
TRANSFORMATION ARTIFACT
MODEL TRAINER CONFIG
MODEL TRAINER
MODEL TRAINER ARTIFACT
EVALUATION CONFIG
MODEL EVALUATION
EVALUATI
IF ACCEPTED
PUSHER CONFIG
MODEL PUSHER
PUSHER ARTIFACT



3.1.2 Prediction Process



3.2 Event Log

The system should log every activity so that user can understand what process is running internally.

3.3 Error Handling

A specific module to be designed to handle all the errors. It should be able to provide specific file name and line number where the error has occured.

3.4 Reusability

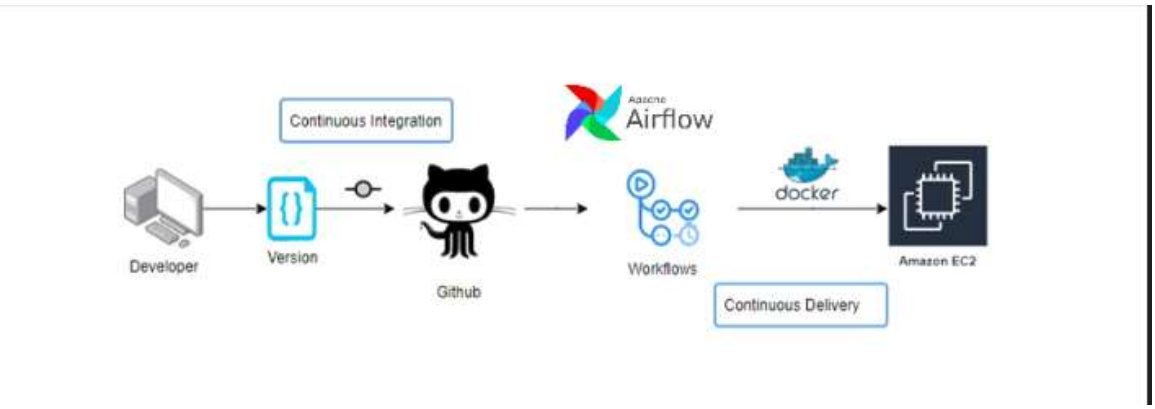
The code written and the components used should have the ability to be reused with no problems.

3.5 Application Compatibility

Each and every components of the project will be using Python as an interface between them and will perform a specific task.

3.6 Deployment

The project is to be delivered using git-Actions workflow and Amazon web Services.



4 Conclusion

The design of this project (thyroid disease detection) will detect the disease have or not the data used to train our algorithm , so we can identify the person have thyroid disease or not ? **We** builded a ML model which will be used by hospitals and help the hospital authority to identify if the patient has thyroid or not. If it is a positive case then medical will do further test to know what type of thyroid the person is suffering from and according to that the treatment will be on fast-track. The doctors will start

treating the patients. If the result will come negative then the patient will be sent to a junior doctor and the junior doctors by using their own expertise they will decide that if the model has done correct prediction or not. If analysis comes true then the doctor release the patient. By seeing the readings if doctor analysed that there may be a chance of thyroid then patient sent to the senior doctors.